

图文 78 生产案例：为什么基于 RocketMQ 进行订单库数据同步时会消息乱序？

403 人次阅读 2020-01-14 09:01:21

详情 评论



狸猫技术

进店逛

相关频道



从 0 开
间件实站
已更新9



继《从零开始带你成为JVM实战高手》后，救火队长携新作再度出山，重磅推荐：

(点击下方蓝字试听)

[《从零开始带你成为MySQL实战优化高手》](#)

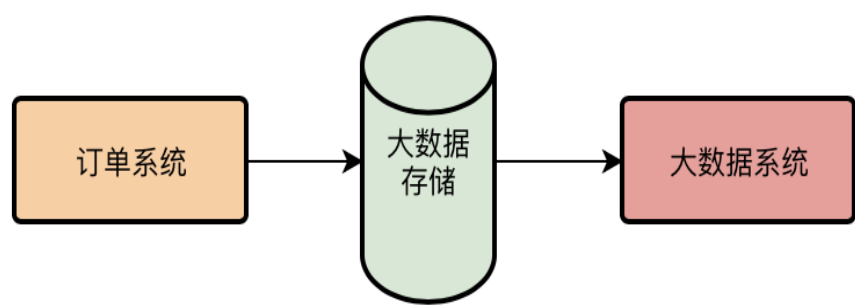
1、大数据团队同步订单数据库的技术方案回顾

在讲完了常规性使用MQ技术的过程中可能遇到的三大问题（消息丢失、消息重复、处理失败）以及其对应的解决方案之后，我们来看一些特殊场景下的MQ的使用问题。

首先我们来看看消息乱序的问题

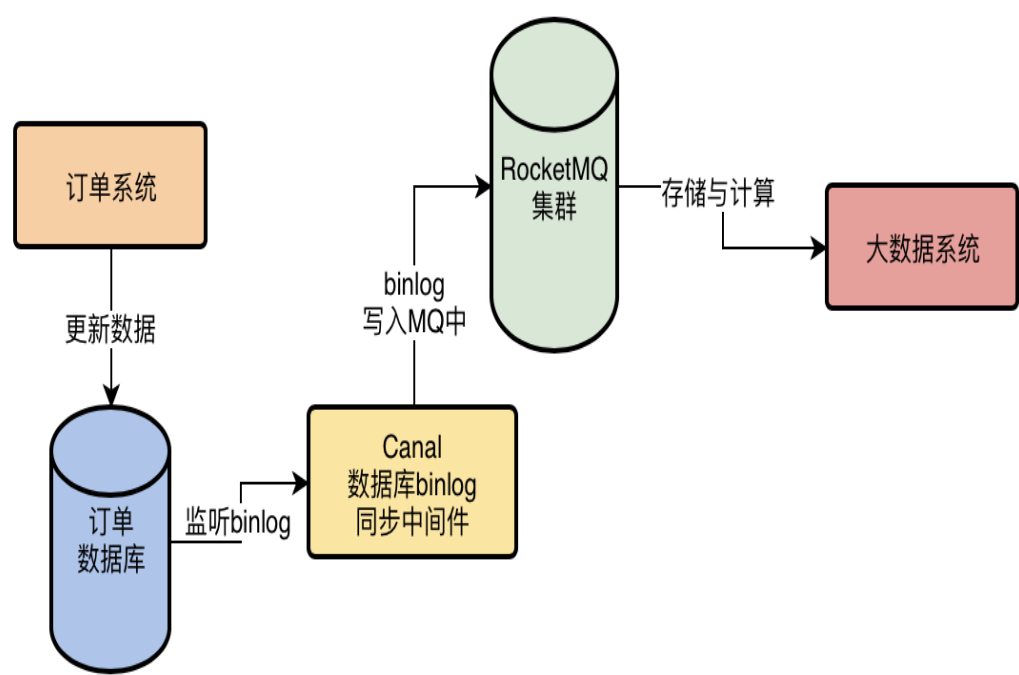
之前我们在讲业务背景的时候，就分析过一个案例，就是大数据团队需要获取订单数据库中的全部数据，然后将订单数据保存一份在自己的大数据存储系统中，比如HDFS、Hive、HBase等

接着基于大数据技术对这些数据进行计算，如下图所示。



然后我们之前讲过，如果让大数据系统自己直接跑复杂的大SQL在订单系统的数据库上来出一些数据报表，是会严重影响订单系统的性能的，所以后来这个方案优化为了，基于Canal这样的中间件去监听订单数据库的binlog，就是一些增删改操作的日志，然后把这些binlog发送到MQ里去。

接着大数据系统自己从MQ里获取binlog，落地到自己的大数据存储中去，然后对自己的存储中的数据进行计算得到数据报表即可，我们看下图。



2、大数据团队遇到的奇怪问题：数据指标错误

这个技术方案原本大家都以为会运行的很良好，结果没想到大数据团队在上了这个技术方案一段时间之后，遇到了一些奇怪的问题。他们通过这个方案计算出来的数据报表，被公司的管理层和运营同事发现，很多数据指标都是错误的！

于是他们就展开了排查，在对自己的大数据存储中的订单数据与订单数据库中的订单数据进行了一次比对之后，发现他们那儿的一些订单数据是不对的。比如在订单数据库中一个订单的字段A的值是100，结果在大数据存储中的一个订单的字段A的值是0。

那如果两边的订单数据的字段值都不一致的话，必然会导致最终计算出来的数据报表的指标是错误的！

3、严密排查之后发现的原因：订单数据库的binlog消息乱序

因此大数据团队针对这个问题，在系统中打印了很多的日志，然后观察了几天，发现了一个惊人的问题，那就是订单数据库的binlog在通过MQ同步的过程中，出现了奇怪的消息乱序的现象！

简单来说，比如订单系统在更新订单数据库的时候，有两条SQL语句：

```
insert into order values(xx, 0)
update order set xxvalue=100 where id=xxx
```

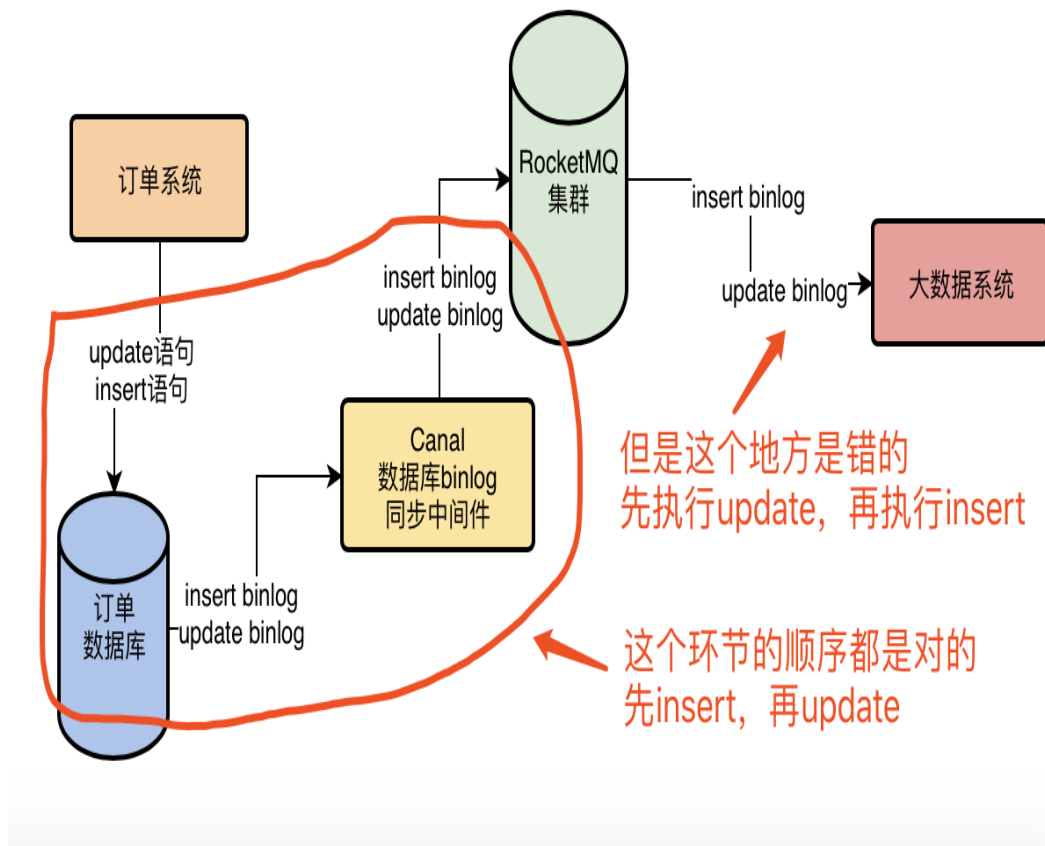
就是先插入了一条订单数据，刚开始他一个字段的值是0，接着更新他的一个字段的值是100。

然后这两条SQL语句是对应着两个binlog的，也就是两个更新日志，一个binlog是insert语句的，一个binlog是update语句的，这个binlog会进入到MQ中去。

然后大数据系统从MQ获取出来binlog的时候，居然是先获取出来了update语句的binlog，然后再获取了insert语句的binlog

也就是说，这个时候会先执行更新操作，但是此时数据根本不存在，没法进行更新，接着执行插入操作，也就是插入一条字段值为0的订单数据进去，最后大数据存储中的订单记录的字段值就是0。

我们看下图，有一个清晰的过程展示。



正是这个消息乱序的原因，导致了大数据存储中的数据都错乱了。

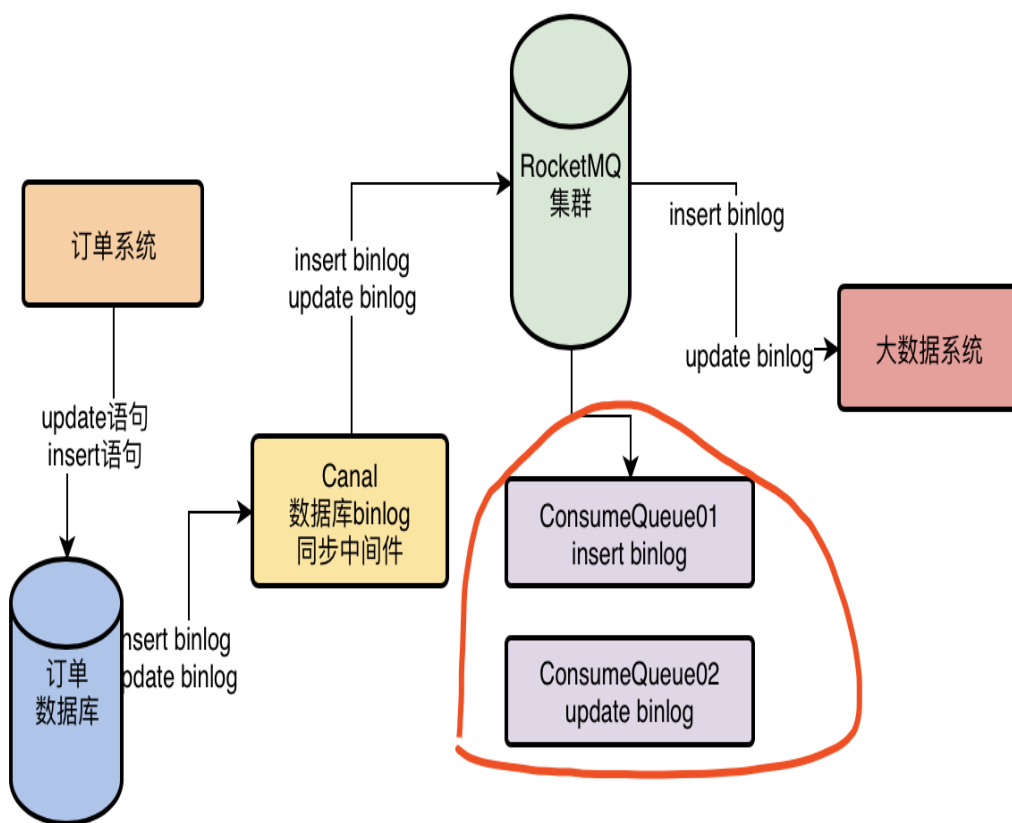
4、为什么基于MQ来传输数据会出现消息乱序？

接着我们来分析一下，为什么基于MQ传输数据的时候默认会出现消息乱序的问题呢？

其实非常简单，我们之前都学习过，可以给每个Topic指定多个MessageQueue，然后你写入消息的时候，其实是会把消息均匀分发给不同的MessageQueue的。

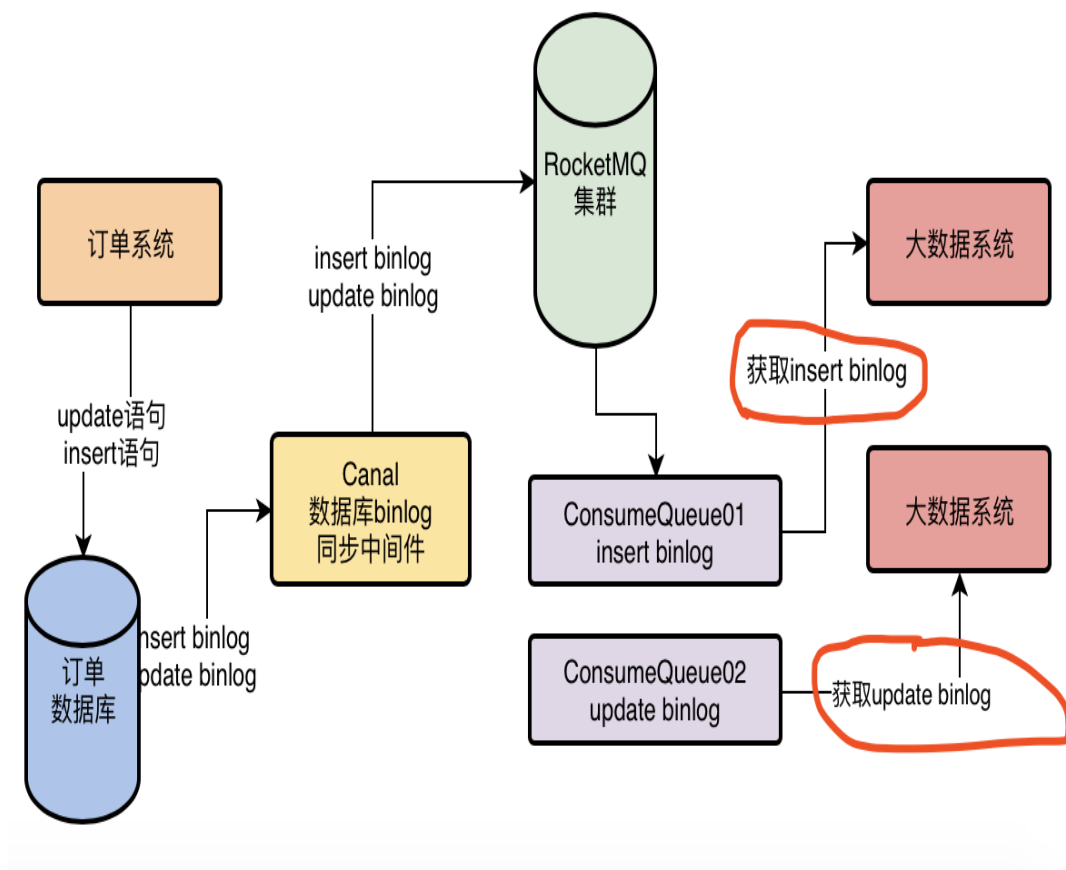
比如我们这里在写入binlog到MQ的时候，可能会把insert binlog写入到一个MessageQueue里去，update binlog写入到另外一个MessageQueue里去

我们看下面的图示



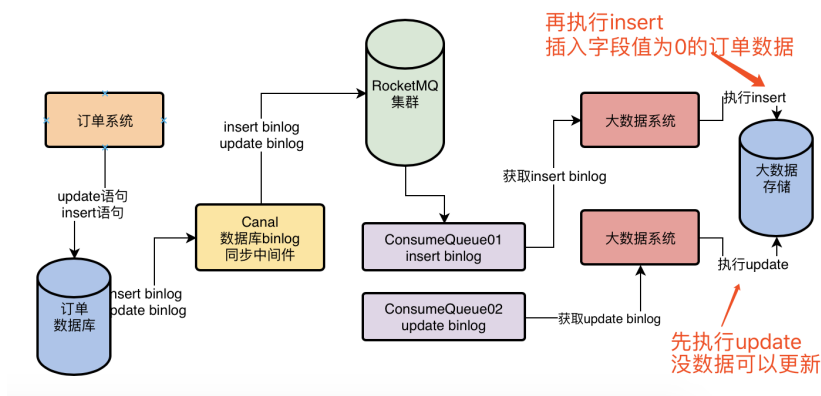
接着大数据系统在获取binlog的时候，可能会部署多台机器组成一个Consumer Group，对于Consumer Group中的每台机器都会负责消费一部分MessageQueue的消息，所以可能一台机器从上图的ConsumeQueue01中获取到了insert binlog，一台机器从上图的ConsumeQueue02中获取到了update binlog

如下图所示



因为我们看到上图中，是两台机器上的大数据系统并行的去获取binlog，所以完全有可能是其中一个大数据系统先获取到了update binlog去执行了更新操作，此时存储中没有数据，自然是没法更新的。

然后另外一个大数据系统再获取到insert binlog去执行插入操作，最终导致只有一个字段值为0的订单数据，如下图：



5、消息乱序：必须要正视的一个问题

所以经过本文的分析，我们完全可以清晰的看到，在使用MQ的时候出现消息乱序是非常正常的一个问题，因为我们原本有顺序的消息，完全可能会分发到不同的MessageQueue中去，然后不同机器上部署的Consumer可能会用混乱的顺序从不同的MessageQueue里获取消息然后处理。

所以在实际使用MQ的时候，我们必须考虑到这个问题。

6、小作业：你们的系统遇到过消息乱序的问题吗？

今天给大家留一个小作业，希望大家思考一下：

你们系统如果使用过MQ的话，到底有没有消息乱序的问题？

如果消息乱序对你们的业务有没有影响？

如果遇到过的话，你们当时是怎么处理的？

希望大家在评论区踊跃的分享自己的经验。

End

专栏版权归公众号**狸猫技术窝**所有

未经许可不得传播，如有侵权将追究法律责任

狸猫技术窝精品专栏及课程推荐：

[《从零开始带你成为JVM实战高手》](#)
[《21天互联网Java进阶面试训练营》（分布式篇）](#)
[《互联网Java工程师面试突击》（第1季）](#)
[《互联网Java工程师面试突击》（第3季）](#)

重要说明：

如何提问：每篇文章都有评论区，大家可以尽情留言提问，我会逐一答疑

如何加群：购买狸猫技术窝专栏的小伙伴都可以加入狸猫技术交流群，一个非常纯粹的技术交流的地方

具体加群方式，请参见目录菜单下的文档：《付费用户如何加群》（**购买后可见**）